1972

A validation of the psychologist's narrative report in predicting "successful" bank managers

Peter Stephen Horos

Follow this and additional works at: http://scholarship.richmond.edu/masters-theses

Part of the Psychology Commons

Recommended Citation
http://scholarship.richmond.edu/masters-theses/1222

This Thesis is brought to you for free and open access by the Student Research at UR Scholarship Repository. It has been accepted for inclusion in Master's Theses by an authorized administrator of UR Scholarship Repository. For more information, please contact scholarshiprepository@richmond.edu.
A VALIDATION OF THE PSYCHOLOGIST'S NARRATIVE REPORT IN PREDICTING "SUCCESSFUL" BANK MANAGERS

BY

PETER STEPHEN HOROS

A THESIS SUBMITTED TO THE GRADUATE FACULTY OF THE UNIVERSITY OF RICHMOND IN CANDIDACY FOR THE DEGREE OF MASTER OF ARTS IN PSYCHOLOGY

AUGUST 1972
A VALIDATION OF THE PSYCHOLOGIST'S NARRATIVE REPORT IN
PREDICTING "SUCCESSFUL" BANK MANAGERS

by

PETER STEPHEN HOROS

Approved:

Robert J. Filer, Chairman
Merton E. Carver
William H. Leftwich
L. James Tromater
# TABLE OF CONTENTS

LIST OF TABLES \hspace{7cm} iii

PREFACE \hspace{7cm} iv

CHAPTER

I \hspace{3cm} INTRODUCTION \hspace{3cm} 1

II \hspace{3cm} METHOD AND PROCEDURE \hspace{3cm} 21

Assessment Procedure \hspace{3cm} 21

Sample \hspace{3cm} 22

Predictors \hspace{3cm} 23

Criteria \hspace{3cm} 24

Procedure \hspace{3cm} 25

III \hspace{3cm} RESULTS \hspace{3cm} 27

IV \hspace{3cm} DISCUSSION \hspace{3cm} 33

V \hspace{3cm} SUMMARY \hspace{3cm} 40

APPENDIX

Appendix A - Rating Procedure \hspace{3cm} 41

BIBLIOGRAPHY \hspace{7cm} 45

VITA \hspace{7cm} 47
**LIST OF TABLES**

**TABLES**

1. Distribution of Predictor Ratings for the Three Banks.  
2. Percentage of Correct Prediction and Deviation from Correct Prediction for the Eight Criterion Measures.  
3. Contingency Coefficients Between the Overall Prediction of "Success" and the Eight Criterion Measures.
To my parents
PREFACE

Recognizing that "no man is an island unto himself" the author would like to express his gratitude to the many people who gave so altruistically of their time, talents, and encouragement during this study. Dr. Robert J. Filer, who chaired the committee, provided both the initiative for the project and continued support throughout its duration. He also provided unlimited access to the reports and facilities necessary to implement these ideas. As the Associate Dean of the Summer School, Dr. William H. Leftwich had numerous responsibilities and duties, but always gave freely of his time and statistical expertise during the many moments of chaos. The author was fortunate to have had the assistance of Dr. Merton E. Carver, retiring chairman of the department of psychology, especially in securing funds for the study. Thanks is also extended to Dr. L. James Tromater, who on that "hectic" final day gave of his time to attend the committee meeting.

A special note of thanks is extended to Mr. James G. Overton, of Psychological Consultants Incorporated. The criterion measures used in the present study came from his earlier research. He was also a source of constant support throughout the study.

A note of thanks to Messrs. Foley, Howie, and Pines, who spent many hours rating the reports. Also to Mr. Joseph
V. Buonassissi, who proofed the statistical calculations and provided constant encouragement. Finally, thanks to all the many people who enabled me to reach this point.
CHAPTER I

INTRODUCTION

Although company selection procedures for managerial personnel vary widely, they frequently contract a psychological consulting firm to assess the applicant's potential. It was generally assumed then, that this professional information would significantly enhance the selection procedures and consequently improve the quality of the organization's managerial population. However, this assumption was made contingent upon the validity of the psychologist's predictions, with which there has been little research to date.

Probably the greatest disparity within this area was with respect to the methodology employed in the psychologist's prediction. Although literature in the 1920's introduced a dichotomy between judgmental and actuarial prediction procedures, it was Meehl's Clinical vs. Statistical Prediction (1954) which provided the arena for this issue. In his book, Meehl defined the parameters for both the kinds of data and the combination procedures necessary for the two polarized approaches. The statistical method utilized psychometric tests which were (1) standardized, (2) objective, and (3) had a reference group (norm); and then combined this data "by some straightforward application of an equation or table (p. 15)." Thus statistical prediction was "derived" from the
data, and because of its mechanical-like qualities could be readily duplicated. "Non-psychometric or case study" data however, was considered to be any information which was not based on psychometric evaluations; and this data was then combined by means of human judgments, "the rules for which are buried in the judges' heads (p. 16)." Consequently, clinical predictions were considered to be "created"from the data.

The remainder of Meehl's book was a systematic presentation and evaluation of 21 studies in an effort to contrast these two methods of prediction. These studies were not directly concerned with predicting managerial success, but dealt rather primarily with college students, prison inmates, and psychotics. The results however, found that in "16 to 20 studies involving a comparison of clinical and actuarial methods, in all but one of which the predictions made actuarially were either approximately equal or superior to those made by a clinician (p. 119)."

In this premier investigation into this area statistical prediction proved to exceed clinical prediction by a substantial margin.

In 1958, the leading proponent of clinical prediction, Robert Holt, published a rebuttal of Meehl's findings. In this article he chastised Meehl for perpetuating competition and controversy by "pitting one method against the other and trying to decide what was the proper sphere of exercise for each (p. 1)."

In examining the foundations for the predictive process, Holt elaborated upon four basic phases (study of the criterion; determining the intervening variables; determining the appropriate measures; and gathering and processing the data) which
occurred prior to the final incorporation of the data. He stated that Meehl's interests were confined mainly to this latter stage. It was from here that Holt's primary criticism evolved, since the two procedures were not comparable during these preliminary stages and consequently could not be adequately contrasted in the final prediction. This was especially evident in the third phase, where the actuarial approach had the advantage of previously "studying their predictive data in relation to the criterion; while the clinician did not (p. 4)."

A further criticism of Meehl's comparisons was Holt's contention that three types of prediction existed: (1) Pure Actuarial - where psychometric data was combined mechanically; (2) Naive Clinical - where assessments were made by clinical judgments based upon qualitative data; and (3) Sophisticated Clinical - where the clinical and actuarial procedures worked in conjunction in an effort to embellish the clinician's reliability and validity, but with the ultimate data organization still residing within the clinician. This latter prediction process exemplified Holt's standards throughout this controversy, since he disavowed any forms of polarity and advocated optimum prediction by means of a combination of the two approaches. However, he felt that Meehl disregarded these possible gradations of clinical prediction and compared Pure Actuarial with an inappropriate clinical approach.

In 1963, Gough presented a historical review of the development of this issue. In his conclusions he reiterated
Holt's concern for caution in the comparisons of these two approaches, especially since the clinical techniques did not have the same validation experiences that the actuarial techniques possessed. He also emphasized that, despite the previous reviews which had reported the actuarial method to have surpassed the clinical, neither approach had been able to report "very good" validity coefficients. Gough therefore, indicated that more accurate predictions were desired and that this might be achieved by incorporating the clinicians skills into the actuarial system.

Sawyer (1966), realized that while clinical and actuarial predictions were being compared, the various methods of data collection were being essentially ignored. Therefore, in his evaluation of forty-five studies he allowed for six different prediction methods. "Data can be collected in three ways (clinically, mechanically, or by both modes) and combined in two ways (clinically or mechanically)." (p. 181) In a further effort to satisfy the requirements of Holt's Sophisticated Clinical approach, he also included two synthesis approaches (clinically and mechanically) where "the prediction of the first stage was permitted to be used as data in the second (p. 183)." For example, in the clinical synthesis the mechanically combined predictions were given to the clinician along with the other data but still allowed him to integrate all the information judgmentally. Again the sample populations were composed primarily of college students, servicemen, mental patients, and prison inmates. Although the predictions did
not deal with managerial success it was the methodology which was of importance. Sawyer's results indicated the mechanical methods of both measurement and prediction to be superior, even when compared with Sophisticated Clinical prediction. His results, however, did not completely disregard the clinical aspect of prediction and consequently found that the clinician could aid in the data collection, but was not advantageous in the combination process.

This issue was still to be resolved. In 1970, Holt again published a criticism of the approaches used in contrasting the two methods of prediction. He reemphasized his five point prediction system (Holt, 1958) and the need for establishing equivalent criterion and measurement foundations prior to comparing the integrated results. Holt, emphasized that both approaches differed with regard to their criterion and measurement techniques, and further stated that appropriate comparisons should examine what the clinicians have been trained to predict and with the techniques they have been trained in, rather than allowing both sides to assess only quantitative data. In this respect Sawyer was admittedly negligent, since only in one half of the 45 studies he reviewed were the measurement and prediction conducted by bonafide clinicians. The remainder of this article then, was a comprehensive attack on Sawyer's conclusions and procedures. In the final analysis, the collection of 45 studies was reduced to five which would be appropriate for comparison.

Certainly the Clinical vs. Statistical controversy can be
extended to encompass managerial selection procedures. In reviewing psychometric and judgmental predictions in this area, Korman (1968) examined the "usual sources". In his article he defined the two methods of prediction in a manner appropriate for this study. Psychometric prediction referred to statistical manipulation of quantified assessment results. Judgmental prediction required an intermediary to integrate the psychometric scores and/or qualitative impressions in some "judgmental" fashion. Korman also stressed the use of absolute level correlations rather than statistical significance. Thus in evaluating some forty managerial selection studies, Korman concluded that the "judgmental prediction methods are generally better predictors than psychometric procedures, although allowances must be made for the generally small samples involved (p. 319)." There have been only a handful of studies conducted in the area for his section on Judgmental prediction included only seven research studies. The results of these studies were positive and generally promising. Consequently it will be with this specific method of prediction, which is probably the most popular though unvalidated approach used today, that the present study will concentrate on.

Two of the earlier studies presented by Korman, were Vernon's 1950 research of Civil Service Managers, and Handy-side and Duncan's 1954 work on first line supervisor personnel. With the former study, the results indicated that although neither approach yielded very high correlations with supervisor ratings, the judgmental method was superior. It was recognized
however, that these applicants constituted a homogeneous, preselected population. Consequently, when the correlations were corrected for range restriction the superiority of the judgmental approach was intensified with correlations now in the .40's and .50's as contrasted with correlations hovering around the .20's for the actuarial prediction. In the latter study, a managerial selection panel had predictive validity coefficients in the .50's and .60's, while tests on verbal and non-verbal abilities correlated .52 and .40 respectively with the promotion rate.

Myer (1956) conducted "an evaluation of a Supervisory Selection Program" used by the General Electric Company. This procedure required Personnel Specialists to prepare summary reports for 139 first line supervisors based upon a combination of psychometric and non-psychometric data. These predictions were correlated with the criterion measures, which were rankings prepared by the second and third level superiors on four dimensions: (1) Human relations; (2) Job knowledge or technical ability; (3) Administrative ability; and (4) Over-all performance.

In evaluating the effectiveness of this selection procedure, Myer found that it was able to predict significantly over-all success when compared with those who had not been evaluated. This conclusion stated the obvious, namely, that any program for selection would be better than none. When Myer examined the validity of the evaluator's overall recommendations he found them to be significantly \( p < .01 \) related to subsequent performance. With respect to the evaluator's ability
to predict the other three characteristics however, the correlations were positive but low, ranging from .11 to .38, with two of them (Administrative Skill and Human Relations Skills) significant at the .05 level. Finally, in examining the specific psychometric measures employed in the evaluation, he found two of the four tests (Wonderlic .27 and Bennett Test of Mechanical Comprehension .29) to be significantly correlated ($p < .01$) with the overall ratings.

In this study, Myer's primary concern was not with a comparison of psychometric and judgmental prediction processes. Thus while his results indicated that the validity of the evaluator's overall prediction was significant, he did not mechanically combine the various psychometric measures, but viewed their predictions separately. The statistical significance reported for the evaluator's overall prediction was determined by applying a Chi Square in comparing the three overall recommendations across the supervisors' ratings of success. Consequently, no absolute level of correlation was computed. Only the validity coefficients for the three specific characteristics could be compared with the previous studies, but no comparisons of overall coefficients can be determined. An additional factor influencing Myer's results was the possibility of criterion contamination, since the supervisors had access to the predictors' judgments in making their criterion ratings.

Campbell, Otis, Liske, and Prien (1962) investigated the validity of predictions made by the psychologists who had all the available information about the person being assessed while
preparing the final reports. The sample populations in this study consisted of two sales and two non-sales groups. Both the psychologists and the first and second level supervisors rated the applicants on eight scales: (1) Social Skills; (2) Persuasiveness; (3) Leadership; (4) Intellectual Capacity; (5) Creativeness; (6) Planning; (7) Motivation and Energy; and (8) Over-All Effectiveness. Correlations were computed between the psychologist's ratings and the two levels of supervisors for the two different populations. The psychologist's predictions were also correlated with the supervisor's ratings on a five step overall effectiveness or "action" criterion scale. Statistical prediction was examined by means of the correlations between the three types of tests used (intelligence, personality inventories, and an interest inventory) and the three criterion scales (creativity, social, and overall).

Appropriate correlations for the psychologist's predictions were, with one exception, positive but low, ranging from -.05 to .50. Correlations with the "action" criterion ranged from .08 to .49. Therefore, the authors concluded that there was some general agreement between the supervisor's evaluations and that of the psychologist. The correlations for the specific test scores with the three criterion scales indicated that the actuarial prediction was less effective.

The authors of the study, never reported statistical significance on their data, and consequently the validity coefficients were all absolute correlations. In defense of these low coefficients however, the authors recognized the
possibility of confounding since several varieties of jobs and raters were contained within the two labels: sales and non-sales. The correlations reportedly improved when the criterion was stated in "action" terms but the authors neglected to define the distinction between these two dimensions adequately. It was also noted that the supervisors tended to rate more leniently and also to rate the sales personnel higher than the non-sales. This implied that the psychologist's predictions were more objective.

Albrecht, Glaser, and Marks (1964) investigated a multiple assessment procedure, which again employed both psychometric and non-psychometric data, but was not utilized in the actual selection process since this was a newly created position. The predictors in this study were assessment reports prepared by a consulting firm which ranked four traits: (1) forecasting-budgeting; (2) sales; (3) interpersonal relationships; and (4) overall performance. Four separate criterion measures included rankings on these same traits by the Regional General Manager; Regional Marketing Manager; Marketing Manager Peers; and ratings by the District Manager. Therefore, there were five method variables and four trait variables.

In evaluating the predictive validity of the assessment reports the authors examined the 16 appropriate correlations (where Pl Sales was correlated with Cl Sales). These correlations ranged from -.04 to .61, between the consultant's rankings and each of the four criterion rankings or ratings. Nine of these validity coefficients were found to be significant at
the .05 level. When the ratings of the District Manager were removed, nine of the 12 coefficients were reported significant. In condensing the number of criterion measures, the three rankings created a composite criterion with the appropriate correlations ranging from .43 to .58, all significant at the .01 level, while none of the inappropriate correlations were larger than .40. Of the eight intercorrelations obtained for the statistical prediction only one was significant. These correlations ranged from -.07 to .41 for the two tests (Problems Test and Watson-Glaser Critical Thinking Appraisal) used. Therefore, it was evident that assessment reports were capable of obtaining superior validity coefficients.

Since this selection procedure was on an a priori basis, none of the applicants rejected for the position appeared in the population surveyed. Consequently, the figures reported here should be corrected for range restriction. In addition the comparison between the psychometric and judgmental prediction procedure should be clarified, since only two tests, both of intellectual ability, were employed. Both tests correlated fairly well with forecasting-budgeting. Perhaps the inclusion of several different tests would have permitted adequate prediction of the other dimensions. Finally, it should be noted that while the intercorrelations between the consultant and the composite rankings were all significant, the correlations between the consultant and the District Manager ratings were all non significant, ranging from .07 to .23. This supported the hypothesis "that the manner in which the criterion is
derived and analyzed may be a very significant factor (p. 359)," and also that the further removed the criterion measure was from the person being assessed the more inaccurate were the impressions.

Miner (1970), who contended that preparing psychological evaluations composed the bulk of activity for the industrial psychologist, conducted a series of seven studies in an effort to examine the validity of these predictions. The first four studies were primarily concerned with non managerial consulting positions and investigated both separatees and present personnel who were employed by both major and minor offices. The appropriate correlations between these psychological evaluations, and the various criterion, rated by the supervisors, were with one exception all non significant and ranged from -.46 to .18. The only significant validity coefficient, -.46; appeared in the fourth study which had a sample population of 24 currently employed personnel in minor offices. Here the psychological evaluations were negatively correlated with the professional grade level, while the correlations with compensation, overall performance, and potential for advancement were all non significant.

The present concern with Miner's study was with the remaining three experiments. In the first, the mean evaluation scores of two managerial groups (upper and lower) were contrasted with those of the separatees and the currently employed major office personnel. This study was based upon the assumption that the evaluations of those presently in managerial positions
should have received generally higher initial evaluations than those who separated without attaining management positions, and also those who were currently employed in non managerial positions. It was further assumed that the evaluations of Upper management would be higher than Lower management. T-tests computed to determine the differences between the psychological evaluations for these groups, reported that none of them were significantly different. The second study was concerned with the tendency for both the individual psychologist and the specific firms to consistently be either too lenient or too harsh. Here the percentage of favorable recommendations and the mean evaluation scores for six psychologists and two firms were compared. The results of the Chi Square indicated that although there was no significant difference between the two firms, there were significant differences between the various psychologists. The third study, which employed the two groups of separatees, was concerned with the fact that various background information may have influenced the psychological evaluations, creating a "Halo" effect. However, this was not supported since none of the correlations were significant.

Although the results of Miner's research contrasted with the previous studies and consequently questioned the predictive validity of the psychological evaluation, it must be recognized that a non managerial sample population was employed. No predictive validity was calculated for the management personnel, and only the mean evaluation scores were compared. Also, in dealing with separatees, Miner was working with a homogeneous
group and consequently these correlations should have been corrected for such restrictions. Finally, in reporting tendencies for particular psychologists to rate either too leniently or too harshly, it appeared that the psychological evaluations became more dependent upon the psychologist involved than any other factors.

All of the preceding studies were primarily concerned with validating the psychologist's final ratings directly with the various criteria. However, frequently these evaluations were in a narrative form which introduced a communication variable not previously considered. In the following studies the final reports were read by individuals other than the report writer, who then quantified the reports with numerical ratings which were correlated with the criterion measures. This was the procedure employed by Walker (1955) in investigating "the effectiveness of communication between the report writer and the business executive reading the report." Although his results indicated that a reasonable proportion of the information was communicable, comparisons between report ratings and final ratings showed a consistent decrease in the validity coefficients. Therefore, this latter section will investigate the validity of the psychologist's final reports indirectly.

Hilton, Bolin, Parker, Taylor, and Walker (1955), in an effort to examine the validity of a Personnel Audit Program, correlated the ratings of report readers with criterion. Their sample population of 100 men was drawn from 18 different companies and 11 different job classifications. The predictors
in this study were independent ratings made by two psychologists, who read each individual's file. This included the test results, interview notes, and the final report. Ratings were then made on five different scales (Sociability; Organizational Ability; Drive; Overall Performance; and Potential). These ratings were correlated with the criterion measures which were supervisor ratings for each man on these same five scales. The results reported that there was general agreement between the supervisor's evaluations and the psychologist's evaluations with the correlations ranging from .21 to .38. Through the use of a multiple factor analysis they also discovered high individual halos both by the predictor and the criterion raters.

Although the validity coefficients in this study were lower than the previous studies, it was difficult to determine whether this was because of the predicted decrease in validity when report ratings were employed or whether the mixed population had an inhibitory effect. It was because of this latter reason that the Hilton, et al. study was not included in Korman's (1968) review of this area. However, although this study utilized report ratings it was not designed to contrast them with final ratings and consequently no legitimate conclusions could be constructed.

Huse (1962), as part of the series of studies conducted by the Western Reserve University, investigated the validity of predictor ratings for the various stages of the prediction process. In order to accomplish this, six sequential assessment
ratings were required: (1) Interview Ratings; (2) Projective Ratings; (3) Test Ratings; (4) Report Ratings - based solely upon reading the final assessment report by a psychologist other than the report writer; (5) Final Ratings - made by the psychologist who wrote the report; (6) Criterion Ratings - consisting of first level supervisors evaluations. Both the predictor and criterion measures then rated eight dimensions: (1) Social Skills; (2) Persuasiveness; (3) Leadership; (4) Intellectual Capacity; (5) Creativeness; (6) Planning; (7) Motivation and Energy; and (8) Overall Effectiveness. Correlations were calculated between the predictor and criterion ratings.

The results revealed that while the predictions based upon Test Ratings were higher than either Interview Ratings or Projective Ratings, they were also equivalent to the Final Ratings. Six of the eight correlations for the Final Rating, while ranging from .13 to .44, were significant (p < .05), in contrast with the Test Ratings, where seven of the eight appropriate correlations were significant and ranged from .15 to .35. When the validity coefficients for the Report Ratings were compared with the Final Ratings there was a consistent decrease in the absolute correlation values, ranging now from .07 to .32. This suggested that while the psychologist was capable of integrating the data and making adequate predictions he had difficulty transferring this information to the report reader and consequently information was lost. Although this study did not support either the psychometric or judgmental prediction procedures, it did indicate that the validity coefficients for
the latter decreased when read by an individual other than the report writer, while the former could still be adequately interpreted.

Dicken and Black (1965), examined the assessment procedures for 31 first line supervisors in a manufacturing firm and 26 in an insurance firm. In this particular program test reports were all written by the same psychologist. They integrated the various test scores and offered a general appraisal of the individual emphasizing his potential for further managerial positions. Both the predictors and criterion measures rated eight global personality variables (Effective Intelligence, Personal Soundness, Drive and Ambition, Leadership, and Dominance, Likeableness, Responsibility and Conscientiousness, Ability to Cooperate, Estimate of Potential Functioning) and two objective measures (Final Salary and Job Level). Predictor data was then obtained through three sources: (1) Report Ratings - where four psychologists read the test reports, (2) Test Scores - purely psychometric approach, (3) Test Ratings (only used with the Insurance Firm) - where to eliminate the effects of the report writer, the predictions were made by a psychologist other than the report writer, who had access to the test data and experience in vocational counseling. The criterion data was secured from (1) Field Ratings - independent raters in the company, (2) Manufacturing Sample Ratings - by second level superiors, (3) Insurance Sample Ratings - by the retired personnel director, current director and vice president, and (4) Objective Criteria - final salary and job level.
This study was based on a three and a half year follow up for the Manufacturing Firm and a seven year follow up for the Insurance Firm.

The validity coefficients for the Manufacturing Firm indicated that five of the eight appropriate correlations between the Report Ratings and Field Ratings were significant \(p < .05\) with correlations from .20 to .51, while only one of the two objective criteria was significant. When corrected for unreliability of the criterion, the correlation values increased from .28 to .57 with seven of the eight now significant. When the correlations from this judgmental approach were contrasted with those from the psychometric procedure, none of the scores, with the exception of one, achieved as high validity as the Report Ratings.

The results for the Insurance Firm were generally not as good, however it must be realized that the population size was smaller and the time lapse had doubled. It was recognized that corrected correlations for the Report Ratings, which ranged from .04 to .66, tended to be smaller than for the Test Ratings, where four of the eight were significant and ranged from .17 to .64. Both predictors failed to predict the objective criterion. Nevertheless, these judgmental predictions were still equivalent to those made by the purely actuarial approach with respect to the personality variables. The results of a matching study conducted with the Insurance Firm, where psychologists were requested to match up test reports with the personnel director's personality sketches, revealed that there was some communality.
since ten of the 12 judges were able to perform this task correctly ($p < .01$).

Although this study did not directly compare the Final Ratings of the psychologist who had written the reports with the Report Ratings, it did reveal that predictions made directly from the data by qualified psychologists were slightly superior to those made following an interpretation of a narrative report. Again, there was a loss of information occurring during the transfer of the actual prediction to the individual reading the report.

The Clinical vs. Statistical Prediction controversy still exists, although the majority of the industrial studies mentioned here, with the exception of Miner (1970), have indicated that the judgmental prediction of managerial success was at least equivalent to the psychometric approach. However, while these studies supported the judgmental side of the controversy they all emphasized the need for additional validation studies in this area. One of the major areas where improvement was necessary was in the selection of appropriate criteria. Only Campbell, et. al. (1962) and Dicken and Black (1965), indicated how their criterion scales were developed, while the remainder of these studies neglected to mention how they arrived at the various measurement scales. Therefore, to predict adequately the individual's success it would be necessary to predict what aspects constituted the successful manager in that area.

Most assessment reports were probably presented to the personnel directors in some type of narrative form. Although
it may have appeared that this allowed the psychologist to de-
velop a global image of the individual he was assessing, the
results of this latter section indicated a decrease in the
validity coefficients the farther removed from the original
prediction one tended to go. Therefore, psychologists in assess-
ment programs must consider how they can communicate their pre-
dictions to the proper personnel effectively.

Finally, the Miner (1970) study investigated the tend-
encies of the individual psychologist to be too lenient or too
harsh, and introduced the possibility that the most important
variable was not any specific data, but which specific psy-
chologist incorporated the data. Consulting firms should be
aware of these tendencies and investigate them further, in an
effort to arrive at a possible weighting system.

The present study is concerned with validating these in-
direct predictions made by sophisticated report readers. There-
fore, it is hypothesized that appropriate correlations, between
the overall evaluations of "success" and the various perfor-
mance measures (job performance, promotability, and salary in-
dex) would be significant. Inappropriate correlations between
overall predictions and five managerial characteristics would
not be significant. In addition, the appropriate validity
coefficients would be superior to previous psychometric predic-
tions made on the same population (Overton, 1971).
CHAPTER II

METHOD AND PROCEDURE

Assessment Procedure

This study was concerned with the predictive validity of the professional psychologist's narrative report, when utilized as a selection technique for bank management trainees. Upon the completion of preliminary screening, three Virginia based banks employed a Richmond consulting firm to further evaluate their applicants. The actual assessment procedure utilized both psychometric and non-psychometric data, which was integrated by a professional psychologist.

The psychometric data included scores on a test battery designed to measure the various attributes believed necessary for a "successful" bank manager. The test battery consisted of all nine of the following tests for banks one and three, and only those indicated by an asterisk for bank two.

1* SRA Verbal
2  Watson-Glaser Critical Thinking Appraisal
3* Guilford-Zimmerman Temperament Survey
4  RBH Vocabulary Test
5  Judgment and Comprehension Test from the Flanagan Aptitude Classification Tests
6  RBH Test of Supervisory Judgment
7  Cardall Arithmetical Reasoning Test
8. How Well Do You Know Your Interests?
9. How Well Do You Know Yourself?

The non-psychometric data included a personal history form and a personal interview. All of this information was then incorporated into a report prepared by the staff psychologist who conducted the interview. The psychologist incorporated the data into a narrative report concerning the applicant's potential for "success" as a bank manager.

After the completion of the assessment procedure, the applicant's written evaluation and test results were forwarded to the respective Personnel Departments. It was there that the final decision as to whether to accept or reject the individual into the training program was made. Therefore, the psychological evaluation was not the sole factor in this selection process, but can be considered an additional and essential tool, used in conjunction with the bank's personal contacts with the individual, and other relevant data.

Sample

The bank management population employed in the present study was the same as the one reported by Overton (1971). This study included 133 present employees who had been initially evaluated by the same Richmond consulting firm. Presently all of these individuals were functioning at some managerial level within the bank. "The sample was restricted to those individuals who had been on the job at least 12 months. A few individuals who had been promoted to top level management positions were not
included because an adequate criterion was not available. Although no records were kept, virtually all of the employees were male Caucasians. A majority of the employees were college graduates (p. 30).

Predictors

In an effort to achieve sophisticated ratings on these global reports, three graduate students in psychology were employed as raters. In a pilot study, conducted by the author, the raters were requested to read independently each report and then rate the individual's overall managerial potential or predictive degree of "success". A five point rating scale, similar to the one used for the criteria, (from Outstanding Potential to Poor Potential) had been devised. The values assigned by the raters were based upon their interpretation of the psychologist's written description of the individual. The results of this study revealed a reliability coefficient of .68 among the three raters. In order to increase this reliability, the raters were trained, through the use of group discussion, to determine on what specific factors their evaluations were established. However, the reliability coefficient remained virtually identical, .66.

In order to avoid using the mean rating it was decided to use a panel discussion. Each report was now read by the raters, who discussed the applicant and his overall potential for "success". The final value was dictated by "majority rule". Twenty reports, selected according to a table of random numbers
(Dixon and Massey, 1957), were retyped and given counterfeit names. These reports were integrated with the original reports and comprised a rating–rerating dimension necessary for the reliability measure.

Criteria

The eight criterion scales employed in this study were developed by Overton (1971). A factor analysis of a 27 item checklist, containing "descriptive statements adapted from the test manuals", revealed five behavioral characteristics. This procedure required the immediate supervisor to rate the employee on these items on a five point scale. Criterion ratings were then determined for the individual on the following five factors:

1. Independence - ability to do the job well with minimal supervision.
2. Interpersonal Relations - ability to get along well with others.
3. Clarity of Communications - ability to transmit and receive information with understanding.
4. Energy and Punctuality - ability to complete work on time without "pushing".
5. Decision-Making Ability Under Pressure - ability to think quickly with good judgment (p. 29).

Overall Performance and Job Promotability criterion scales were developed by means of a forced choice distribution. Supervisor panels were assembled by the Personnel Director, for the respective banks, and the panel members were requested to sort the employees into one of five distributions (10%, 20%, 40%, 20%, 10%), ranging from poorest to best. For the former criterion the panel was asked the following question: "Considering all
factors, where does this employee rank in relation to other workers in terms of his on-the-job performance and competence in his present job (not how well you like him, but how good a job he's doing for the bank)"

For the latter criterion, the panel members were asked: "Where does this employee rank in terms of his promotability to jobs of higher responsibility?"

Finally, the employee's monthly economic growth was measured by a salary index. This "relatively" objective criterion was measured by a statistic developed by Overton (1971).

\[
\text{Salary Index} = \frac{\text{Present Monthly Salary} - \text{Initial Monthly Salary}}{\text{Length of Service in Months}}
\]

In order to eliminate the various sources of financial discrepancy between banks, these ratings were transformed into standard scores with a mean of 10 and a standard deviation of one.

**Procedure**

The narrative reports were to be quantified by the raters. Eight Pearson Product Moment correlations were to be computed between the overall predictor rating and the eight criterion scales. Correlations between the prediction of overall job "success" and various performance criteria (job performance, promotability, and salary index) were to be considered "appropriate" correlations. Correlations between overall job "success" and the five managerial characteristics were to be considered "inappropriate" correlations. All Pearson r's were to be corrected for errors of grouping (Guilford, 1965), since the sample
population was homogeneous, containing only "present" employees. Coefficients exceeding the .05 level were to be considered significant. The magnitude and level of significance of these eight validity coefficients, which were based upon the "judgmental" prediction scheme, were to be compared with those found in the literature. These correlations would also be compared with the multiple correlations revealed by Overton (1971).
CHAPTER III

RESULTS

The Pearson Product Moment correlation computed on the 20 counterfeit reports composing the rating - rerating dimension, yielded a reliability coefficient of $r = .75$. When corrected for errors of grouping (Guilford, 1965) the coefficient was increased to $r = .81$. Tiffin and McCormick (1965, p. 255) reported that the reliability of ratings was contingent upon the type of rating method employed. They reported that reliability coefficients were typically lowest for the rating scale method and that their results yielded an $r = .55$ for this particular method.

The Pearson $r$'s for the overall predictor rating and the various criterion variables were not calculated due to the lack of variability among the report ratings. As illustrated in Table 1, 76 percent of all the ratings were deposited into only two of the five designated categories. Despite this impediment, the percentage of agreement between predictor and criterion scores was tabulated and is reported in Table 2. All computations were based on a population size of 133, with the exception of salary index where the population size was
Table 1

Distribution of Predictor Ratings for the Three Banks

<table>
<thead>
<tr>
<th>Bank</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0</td>
<td>1</td>
<td>13</td>
<td>6</td>
<td>3</td>
</tr>
<tr>
<td>2</td>
<td>0</td>
<td>1</td>
<td>10</td>
<td>15</td>
<td>2</td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>13</td>
<td>40</td>
<td>18</td>
<td>10</td>
</tr>
<tr>
<td>Total</td>
<td>1</td>
<td>15</td>
<td>63</td>
<td>39</td>
<td>15</td>
</tr>
<tr>
<td>Per Cent</td>
<td>0%</td>
<td>11%</td>
<td>47%</td>
<td>29%</td>
<td>11%</td>
</tr>
</tbody>
</table>
Percentages were found to range from 23 to 45 for correct matches, increasing to 70 to 80 when those predictors deviating one rating category from the criterion value were included.

Insert Table 2 about here

In further analyzing the data it was necessary to collapse some of the categories because of poor rater variability. The Contingency Coefficient (C) was selected because it did not require underlying continuity in the variables under analysis and made no assumptions concerning the shape of the population of scores. In performing the collapsing operation the basic assumptions of the Chi Square had to be met, since this was a basic component of the C-Coefficient formula. The final scheme collapsed categories four and five on the criterion dimension and categories one and two and four and five on the predictor scale. Only two of the eight criteria however, were based on a forced distribution rating scale from one to five, while the other six were expressed in standard scores. Consequently, in an effort to establish a corresponding distribution (10%, 20%, 40%, 20%, 10%), z-score terminating points were calculated (-1.282, -.524, +.524, +1.282). By collapsing the data in this fashion (1) all those cells containing a value of "zero" were eliminated and (2) there were less than 20 percent of the cells containing values of "one". Consequently, the basic assumptions of Chi Square were met.

The C-Coefficient has a number of inherent limitations in
Table 2

Percentage of Correct Prediction and Deviation from Correct Prediction for the Eight Criterion Measures

<table>
<thead>
<tr>
<th>Criteria</th>
<th>Deviation from Prediction</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0</td>
</tr>
<tr>
<td>Job Performance</td>
<td>30</td>
</tr>
<tr>
<td>Promotability</td>
<td>29</td>
</tr>
<tr>
<td>Salary Index</td>
<td>45</td>
</tr>
<tr>
<td>Independence</td>
<td>27</td>
</tr>
<tr>
<td>Interpersonal Relations</td>
<td>33</td>
</tr>
<tr>
<td>Clarity of Communications</td>
<td>29</td>
</tr>
<tr>
<td>Energy</td>
<td>30</td>
</tr>
<tr>
<td>Decision Making</td>
<td>23</td>
</tr>
</tbody>
</table>
its applicability but probably the most relevant to the present study was the fact that it could not be appropriately compared with any other measures of correlation. Furthermore, while the C-Coefficient will yield a "zero" coefficient when there is no agreement between two variables, it will not attain a value of unity, but only approach it. Therefore, an upper limit must be determined for the C-Coefficient, which was dependent upon the number of categories utilized. The maximum C in this study was found to be .87 when computed for four categories according to Siegle (1956, p. 201).

The C-Coefficients for the overall prediction with the eight criterion measures are reported in Table 3. The C-Coefficients for the three indexes of performance (job performance, promotability, and salary index) were .31, .32, and .37 respectively and were all significant (p < .05). The coefficients for the five managerial characteristics meanwhile ranged from .15 to .38, with only Clarity of Communications attaining significance. Therefore, four of the eight Contingency Coefficients are reported as significant.

Insert Table 3 about here
Table 3

Contingency Coefficients Between the Overall Prediction of "Success" and the Eight Criterion Measures

<table>
<thead>
<tr>
<th>Criteria</th>
<th>C-Coefficients</th>
</tr>
</thead>
<tbody>
<tr>
<td>Job Performance</td>
<td>.31 *</td>
</tr>
<tr>
<td>Promotability</td>
<td>.32 *</td>
</tr>
<tr>
<td>Salary Index</td>
<td>.37 *</td>
</tr>
<tr>
<td>Independence</td>
<td>.24</td>
</tr>
<tr>
<td>Interpersonal Relations</td>
<td>.23</td>
</tr>
<tr>
<td>Clarity of Communications</td>
<td>.38 *</td>
</tr>
<tr>
<td>Energy and Punctuality</td>
<td>.15</td>
</tr>
<tr>
<td>Decision Making</td>
<td>.29</td>
</tr>
</tbody>
</table>

* p < .05
CHAPTER IV

DISCUSSION

It was necessary to artificially collapse both the predictor and criterion scales because of the lack of variability among the predictor ratings. The criterion scale imposed forced variability which was normal in form and based on a continuum ranging from the poorest to the best. The predictor scale utilized the same continuum and found that 76 per cent of the reports appeared in the "average" and "above average" categories. This indicated that the lower extreme of the continuum was being neglected. This was understandable considering the selective population (only current employees) and recognizing that the raters were not required to force the distribution. Consequently the use of a Product Moment correlation would have concealed any degree of relationship between these two variables.

A miscalculation in the criterion scale for "job performance" also created a need to re-examine the continuity of the underlying variables. Performance appraisals conducted by the Personnel Directors were forced in a normal distribution (10%, 20%, 40%, 20%, 10%). On this particular dimension one bank had a positively skewed distribution (10%, 40%, 20%, 20%, 10%). It was unknown how or why this occurred. For this reason, it was believed the predictor and criterion scales could not be compared.
All of these factors contributed to the need to collapse the number of categories. A non-parametric statistic was selected because of the questions raised about the continuity of the underlying categories and the distribution of the selective population. The Contingency Coefficient which was free from these assumptions and requirements, premitted the collapsing of various categories while measuring the association between the two variables. The major consequence of employing this statistic was that the C-Coefficients and the Product Moment coefficients could not be compared. It was also impossible to correct the C-Coefficient for errors due to grouping. This eliminated the possibility of comparing the correlation coefficients of this study with those previously reported in the literature. Levels of significance were still able to be contrasted.

C-Coefficients were calculated between the predictions of overall success, based on the narrative reports, and the eight criterion variables. Table 3 indicated that four of the eight correlations were significant \( (p < .05) \). In examining "job performance" \( (C = .31) \), "promotability" \( (C = .32) \), and "salary index" \( (C = .37) \) it was observed that all three were based upon actual performance and managerial potential. They were considered appropriate criterion for predictions of overall performance. "Performance" and "salary index" were measured by the supervisors evaluation of job performance and more objectively the individual's economic growth rate, while "promotability" was an estimate of the individual's potential to assume greater job responsibility. All three coefficients were
significant, indicating that the overall prediction of job success was directly related to future job performance and potential. Additional support for this conclusion was reported in Table 1. There the percentages for correct matches and deviations from perfect predictions were reported. "Salary index" which attained the highest C-Coefficient of the three was predicted within one category for 80 per cent of the cases. Predictions of overall success were capable of predicting "job performance" and "promotability", within the same boundaries, for 72 and 76 per cent of the cases respectively.

Although all three of these coefficients were significant, "job performance" resulted in the smallest correlation coefficient. This was probably influenced by the positively skewed distribution which one bank inadvertently employed. Meanwhile, the overall performance correlation with "salary index" reported the best coefficient. This prediction of the individual's economic growth rate was not subjected to supervisor evaluations. The calculation was "relatively" objective by means of the statistical technique developed by Overton (1971).

\[
\text{Salary Index} = \frac{\text{Present Monthly Salary} - \text{Initial Monthly Salary}}{\text{Length of Service in Months}}
\]

Realizing the necessity for adequately measured criteria this formula eliminated some of the effects of interbank differences and economic prosperity prevalent during employment. The strength of the correlation was partially reflected by the effectiveness of this criterion measure.
The success of the overall predictions of performance when correlated with these three appropriate criteria was hypothesized and well established in the literature (Hilton, et. al., 1955; Huse, 1962; Campbell, et. al., 1962; and Albrecht, et. al., 1964). Only the Miner (1970) study denied this relationship. The inappropriate correlations, where overall performance was used to predict the five managerial characteristics, were not expected to exhibit the same degree of association. This was supported by the results reported in Table 3. Only one of the five correlations, Clarity of Communications ($C = .38$), was significant. Although the significance of this characteristic suggested a possible emphasis on the applicants ability to communicate during the interview, it was important to realize that this factor loaded on only one item. The literature then, supported the present results. Studies by Hilton, et. al. (1955) and Albrecht, et. al. (1964) reported appropriate correlations between the various managerial characteristics to be significant, while inappropriate correlations between the overall performance ratings and these specific characteristics were typically not significant.

Hilton, et. al. (1955), using a mixed population, reported appropriate correlations between the prediction of the overall success and actual job performance to be significant ($p < .05$) with $r = .29$. Appropriate correlations with four managerial characteristics (Sociability, Organizational Ability, Drive, and Advancement Potential) were also significant. However, only two of the four inappropriate correlations between overall per-
formance and these four characteristics were significant. One of these was the prediction of advancement potential, similar to the present "promotability", with \( r = .34 \).

Albrecht, et. al. (1964) substantiated these conclusions and extended them to a specific managerial population. The appropriate correlation for overall performance was significant \( (p < .01) \) with \( r = .46 \). The appropriate correlations for the three managerial characteristics (forecasting, sales, interpersonal relations) were also significant. None of the inappropriate correlations however, exceeded the coefficients for the appropriate correlations. Possible explanations for the discrepancy in the Product Moment coefficients for these two studies was that in the latter direct predictions were made by the consultant while in the former the predictions were based on the "report reader's" interpretation of the narrative report. In addition the Albrecht, et. al. article employed a homogeneous managerial population as opposed to the "mixed" population used by Hilton, et. al.

Unfortunately, no direct comparisons were able to be made between the present validity coefficients for judgmental prediction and those found in the literature for statistical prediction because of the different techniques used to determine the degree of relationship. However Pearson Product Moment correlations between numerous test scales and criteria have been reported below the correlations for the judgmental predictions. Huse,(1962); Campbell, et. al. (1962); and Albrecht, et. al. (1964) all reported statistical prediction to be less
effective. Overton (1971) reported seven, eight, and nine test scales correlated significantly with the 93 total scales for performance, promotability and salary index respectively. The coefficients for these three criteria ranged from -.34 to +.34.

By extending the definition of "statistical" prediction Overton (1971) mathematically combined the prediction afforded by the various psychometric measures. This was the only study found in the present survey of industrial literature which computed a multiple correlation. The validity coefficients for performance, promotability, and salary index were .43, .659, and .516 respectively, overshadowing most of the previously cited studies which considered these predictions separately. The results were not surprising since the simple Pearson r's used only one predictor and could not be compared to the multiple correlation which used several predictors.

The present results supported the contention that predictions of overall success (made by sophisticated raters) based on narrative reports were indicative of future job performance and potential. For future research it might be hypothesized that the validity coefficients will be higher if the predictions are made directly by the psychologist. Although four of the five coefficients for the managerial characteristics were not significant this was expected, and did not cast any suspicions on the inclusion of these elements into the composite of the "successful" bank manager. The narrative reports were written without a knowledge of these factors and
the raters were not required to predict these specific characteristics. Further study will have to be conducted before any conclusions can be made concerning these characteristics. The course of future research should also investigate the application of the multiple correlation scheme in the continued effort to improve managerial prediction.
CHAPTER V

SUMMARY

The purpose of the present study was to investigate the predictive validity of the narrative report used to assess potentially "successful" bank managers. All 133 reports were prepared by the psychologists of a local consulting firm and were based on psychometric and non-psychometric data. Three sophisticated "report readers" independently read each report and discussed the individual's qualifications prior to assigning a rating of overall potential for "success". This one predictor value was then correlated with three appropriate performance measures (job performance, promotability, and salary index) and five managerial characteristics (independence, interpersonal relations, clarity of communications, energy and punctuality, and decision-making ability under pressure).

The results supported the hypothesis that predictions of overall success based on these narrative reports were related to job performance and potential. All three Contingency Coefficients for the appropriate performance measures were significant ($p < .05$). Predictions of overall success were significantly correlated with one of the five managerial characteristics. The results alluded to the possibility that improved predictions could be attained through direct predictions made by the psychologist and the use of multiple correlations.
APPENDIX A

RATING PROCEDURES
This study is concerned with the validation of a selection technique involving Bank Management positions. Upon completion of preliminary screening, three Virginia based Banks utilized an outside consulting firm to further evaluate their applicants. Written reports, assessing the individual's global potential, were prepared by professional psychologists on the basis of a battery of nine tests and occasionally an interview. The applicant's written evaluation was then returned to the Bank's Personnel Department where the final decision to accept the individual into their program was based upon these written reports, the Bank's contacts with the applicant, and other relevant factors.

In your capacity as a Rater you will be requested to read each of the following written reports. Upon reading each report you will be required to rate the individual's overall managerial potential or predictive degree of "success". The value you assign should be based upon your evaluation of the psychologists written description of the individual. In order to facilitate reliable ratings it will be helpful to mention some of the possible "pitfalls" involved in the rating procedure, since research has illustrated that exposure to these problems will eliminate additional sources of variance.

1) Central Tendency

This occurs when the rater neglects to utilize the extreme judgments and confines his ratings to the central values. To avoid this error ratings should be normally distributed throughout the scale.
2) Leniency and Harshness

This is the tendency for some rater to be consistently "hard" and give unfavorable ratings, while others are "easy" and give favorable values. Again, ratings should be normally distributed.

3) Halo and Pitch Fork Effect

Tendency for the rater to be influenced by some specific trait to the extent that it is generalized to the individual's overall ability rating. To avoid this error, ratings should be based upon all factors and not merely dominated by a single element.
"In evaluating the global potential for managerial "success" this individual has

5. OUTSTANDING POTENTIAL:
   Among the best candidates for "success".

4. ABOVE AVERAGE POTENTIAL:
   But not outstanding chance for "success".

3. AVERAGE POTENTIAL:
   Neither above average nor inferior chance for "success".

2. BELOW AVERAGE POTENTIAL:
   But not poor chance for "success".

1. POOR POTENTIAL:
   Falls among the poorest candidate for "success".

(Circle the appropriate number)


